Prof. Philip Koopman

# Adventures in Self Driving Car Safety

March 3, 2020

Carnegie Mellon University

@PhilKoopman

EDGE CASE RESEARCH

# Preview of Adventures

Carnegie Mellon University

- **Autonomous vehicle safety**
  - The hard part is perception/prediction
  - Automated identification of perception mistakes


- **Vehicle Safety**
  - Unintended Acceleration & the pedal misapplication narrative
  - UL 4600: a safety standard for self-driving cars

*Also, personal experiences being an agent of change*

# Autonomy 98% Solved For 25 Years

Carnegie Mellon University


- **Washington DC to San Diego**
  - CMU Navlab 5
  - Dean Pomerleau
  - Todd Jochem

  https://www.cs.cmu.edu/~tjochem/nhaa/nhaa_home_page.html
- **AHS San Diego demo Aug 1997**


# Before Autonomy Software Safety

Carnegie Mellon University

- The Big Red Button era



# APD (Autonomous Platform Demonstrator)

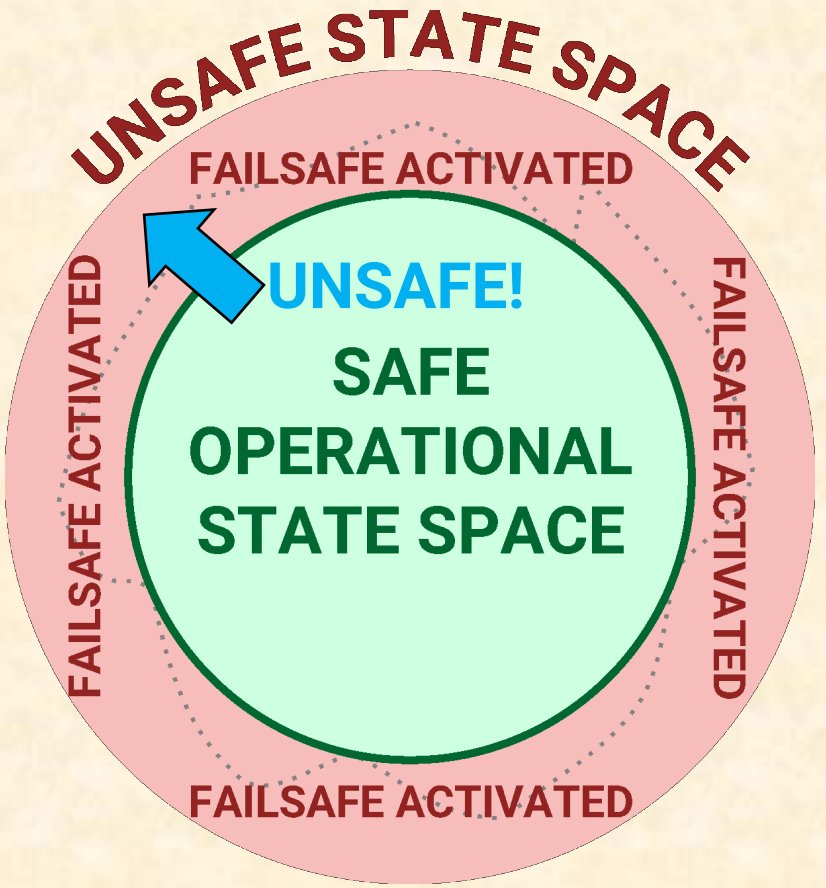
Carnegie Mellon University


NATIONAL ROBOTICS NREC ENGINEERING CENTER

# Safety Envelope Approach to ML Deployment

Carnegie Mellon University

- **Specify unsafe regions**
- **Specify safe regions**
  - Under-approximate to simplify
- **Trigger system safety response upon transition to unsafe region**



- **RSS Following Distance Equation**
  - **Intel/ Mobileye**


Figure 1. Reference vehicle geometry for leader/follower.

This yields a minimum following distance (id., Lemma 2):

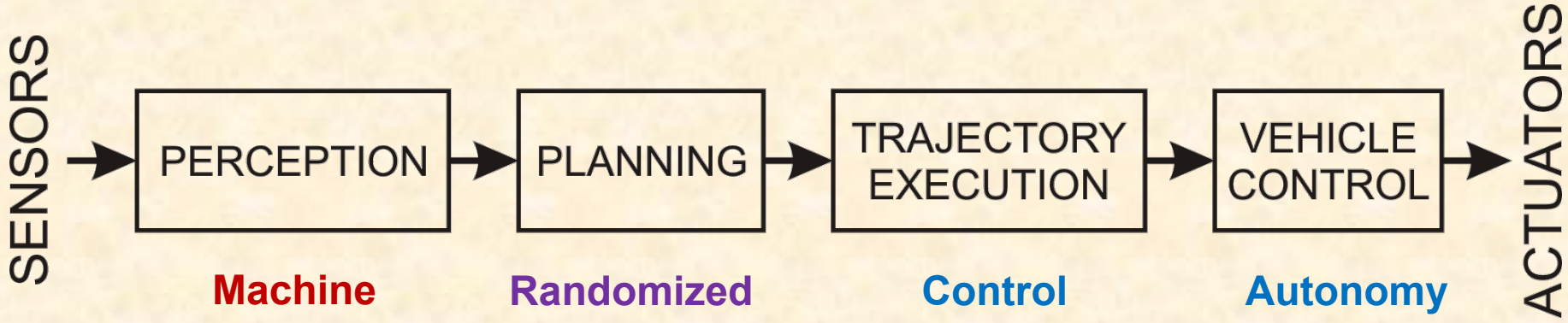
$$d'_{min} = MAX\left\{0, \left(v_r\rho + \frac{1}{2}a_{max,accel}\rho^2 + \frac{(v_r + \rho a_{max,accel})^2}{2a_{min,brake}} - \frac{v_f^2}{2a_{max,brake}}\right)\right\} \quad (1)$$

Where in our case the ego vehicle is the following ("rear") vehicle, and:

- $d'_{min}$ is the minimum following distance between the two vehicles for RSS
- $v_f$ is the longitudinal velocity of the lead ("front") vehicle
- $v_r$ is the longitudinal velocity of the following ("rear") vehicle
- $\rho$ is the response time delay before the ego (rear) vehicle starts braking
- $a_{max,brake}$ is the maximum braking capability of the front vehicle
- $a_{max,accel}$ is the maximum acceleration of the ego (rear) vehicle
- $a_{min,brake}$ is the minimum braking capability of the ego (rear) vehicle

# Validating an Autonomous Vehicle Pipeline

EDGE CASE RESEARCH




**Machine Learning Based Approaches**

➔ ???

**Randomized & Heuristic Algorithms**

➔ Run-Time Safety Envelopes (~don't hit stuff)

**Control Systems**

➔ Control Software Validation

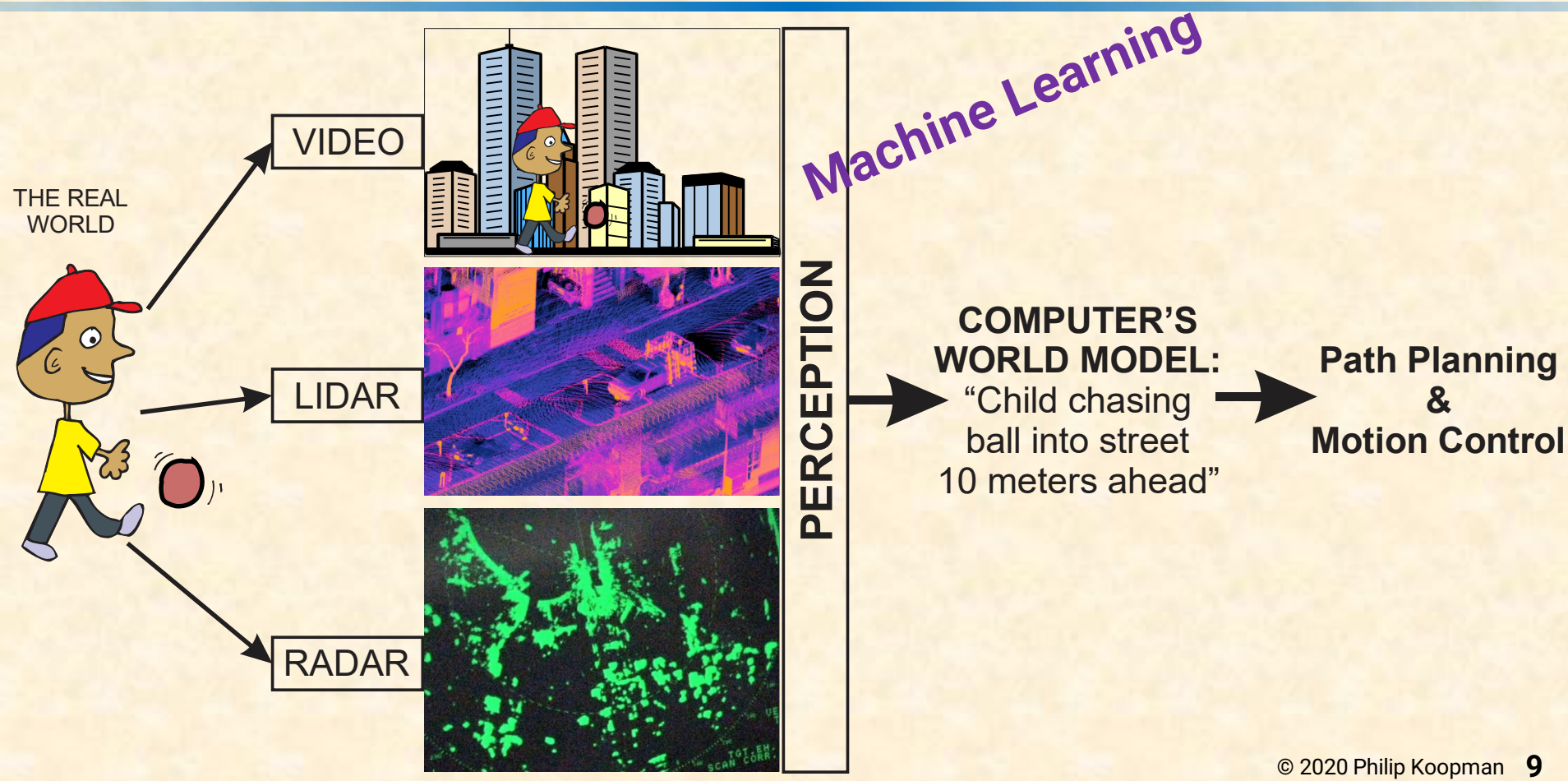
**Autonomy Interface To Vehicle**

➔ Traditional Software Validation

**ML for perception/prediction is uniquely difficult to assure**

# Perception Builds the World Model

Carnegie Mellon University



# Importance of Behavior Prediction

Carnegie Mellon University

- **Free space: available drivable area**
  - Move to where the free space is going to be
  - Requires fine grain classification


# Perception via Machine Learning

EDGE CASE RESEARCH

- **Machine learning uses training data**
  - "Learns" via visual features in a picture

**The proverbial Black swan**

https://bit.ly/3a2cFL7

# Brute Force Road Testing

EDGE CASE RESEARCH

- **Good for identifying "easy" cases**
  - Expensive and potentially dangerous

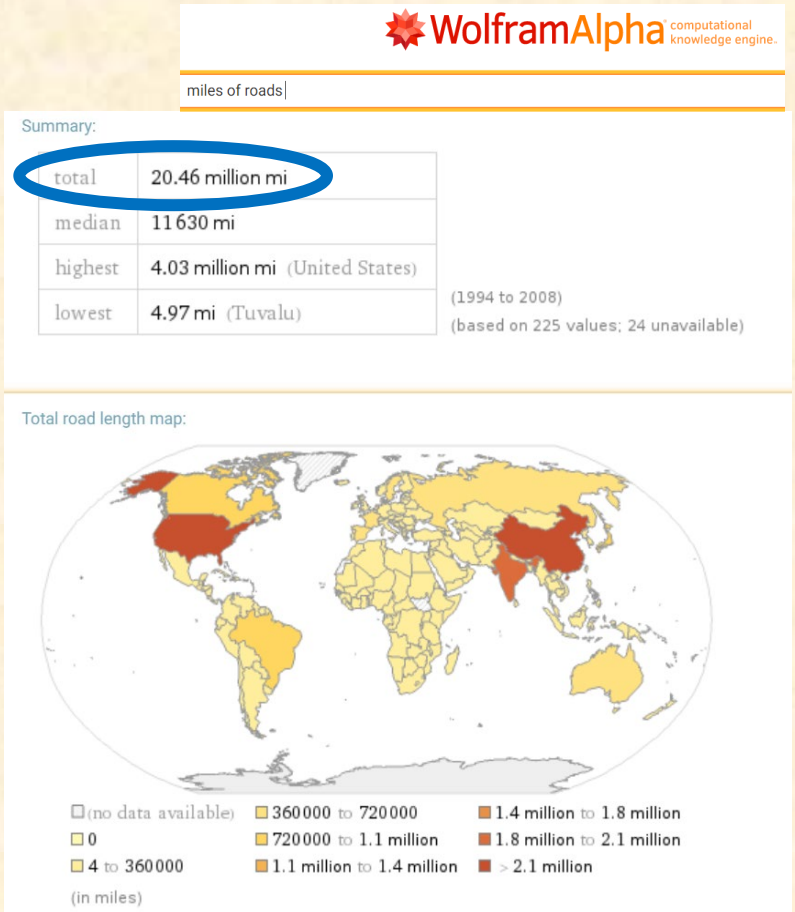

 

# Unrealistic to Brute Force Safety

EDGE CASE RESEARCH

- **If 100M miles/fatal mishap…**
  - Test 3x–10x longer than mishap rate
    ➔ Need 1 Billion miles of testing

- **That's ~25 round trips on every road in the world**
  - With fewer than 10 critical mishaps



# Closed Course Testing

EDGE CASE RESEARCH

- **Safer than road testing, but not scalable**
  - Simulation is scalable
- **But – only test things you have thought of!**



Volvo / Motor Trend

# It's All About The Edge Cases

EDGE CASE RESEARCH

- **Gaps in training data can lead to perception failure**
  - Safety needs to know: "Is that a crossing pedestrian?"
  - Machine learning provides: "Does that look like the crossing pedestrians in training data?"

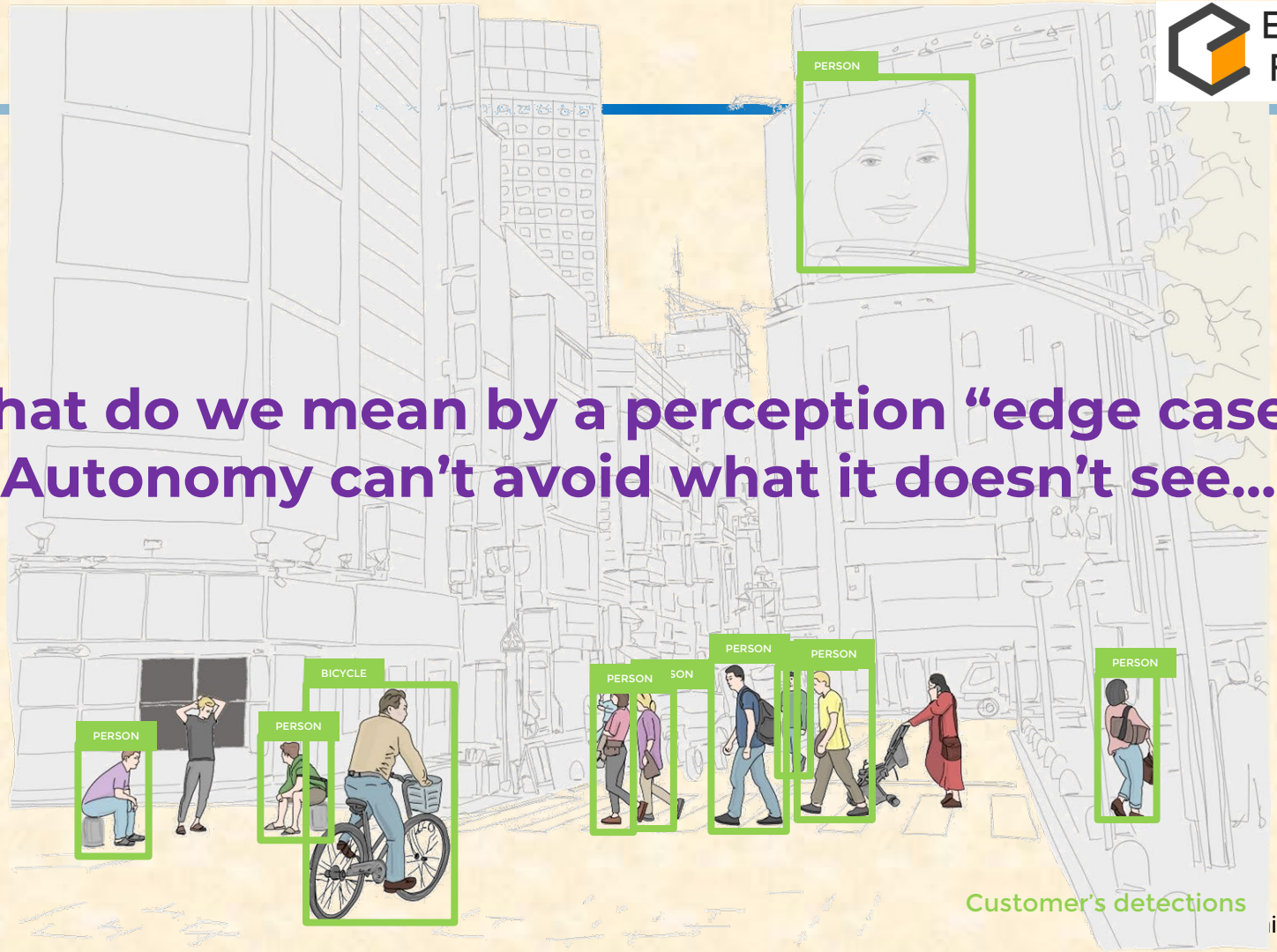

| PREDICTED CONCEPT | PROBABILITY |
|---|---|
| bird | 0.997 |
| no person | 0.990 |
| one | 0.975 |
| feather | 0.970 |
| nature | 0.963 |
| poultry | 0.954 |
| outdoors | 0.936 |
| color | 0.910 |
| animal | 0.908 |

https://www.clarifai.com/demo

- **Edge Case are surprises**
  - ➔ Edge cases are the stuff you didn't think of!

EDGE CASE RESEARCH

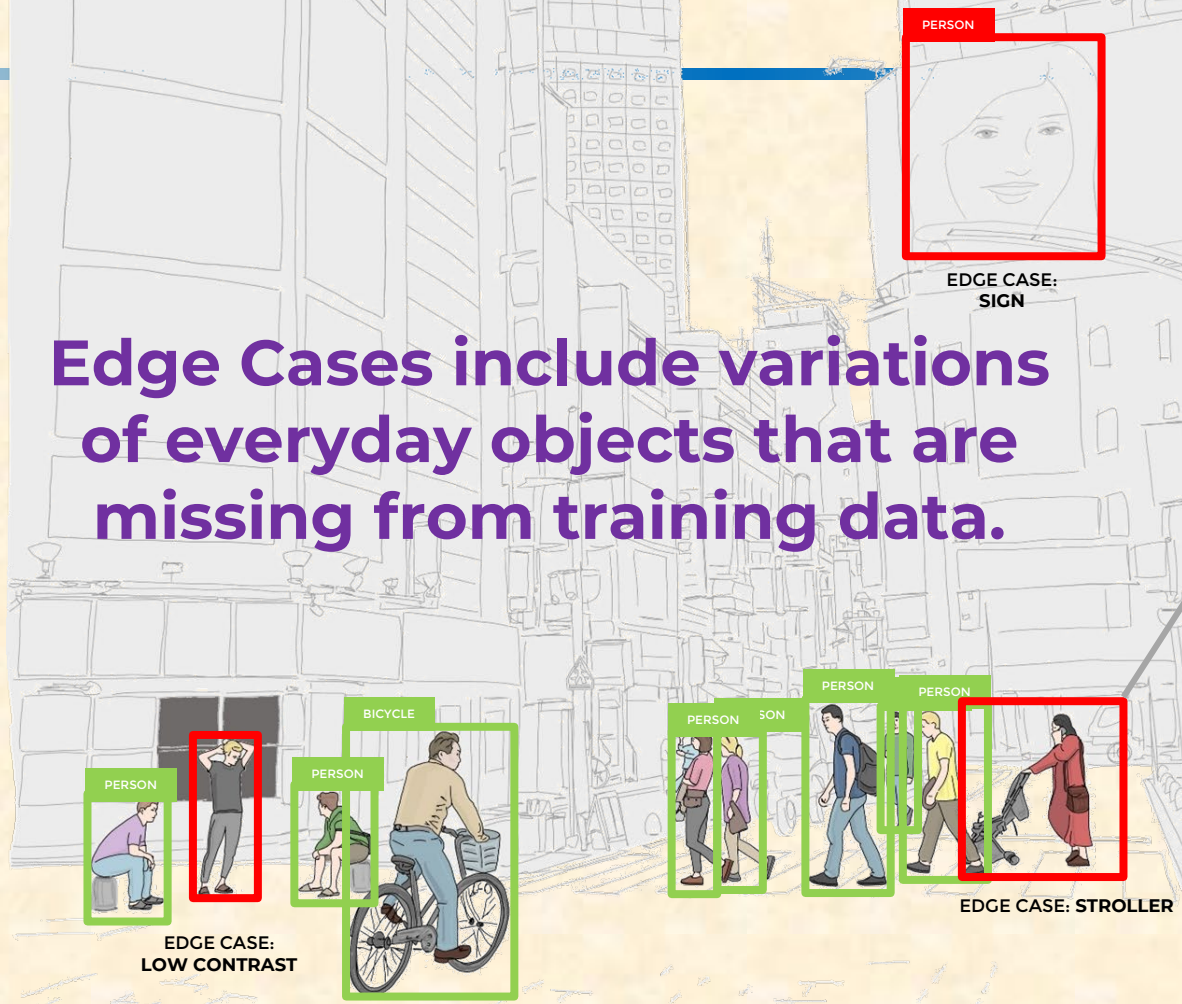
# What do we mean by a perception “edge case”?
# Autonomy can’t avoid what it doesn’t see…




Customer’s detections

EDGE CASE RESEARCH

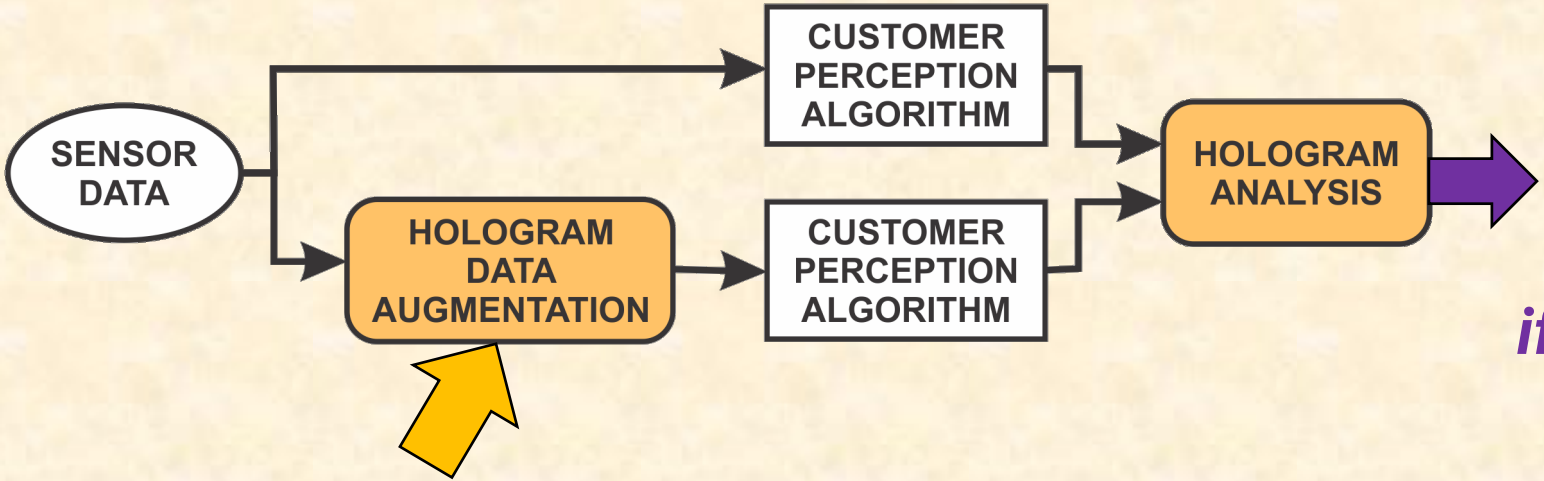
# Edge Cases include variations of everyday objects that are missing from training data.



# Automatically Detecting Edge Cases

EDGE CASE RESEARCH

- **Adding noise to an image causes objects to drop out**
  - Reveals systematic perception issues on unlabeled data




**Simplified example: add light Gaussian Noise**

HOLOGRAM

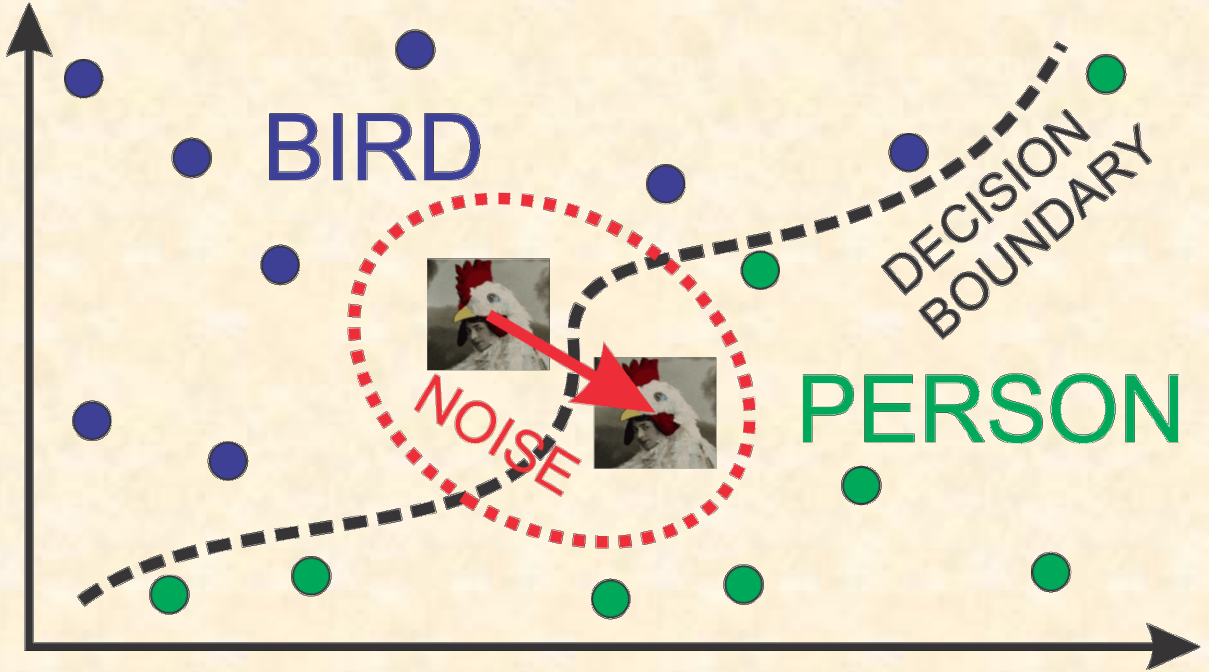
# Pedestrian False Negative

EDGE CASE RESEARCH

# Finding Edge Cases Experimentally

EDGE CASE RESEARCH

- **Noise randomly perturbs data sample in decision space**
  - Change in classification detects points near decision boundary
  - Many of these are Edge Cases (unknown unknowns)

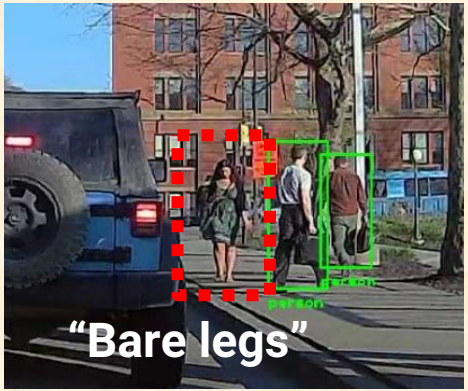


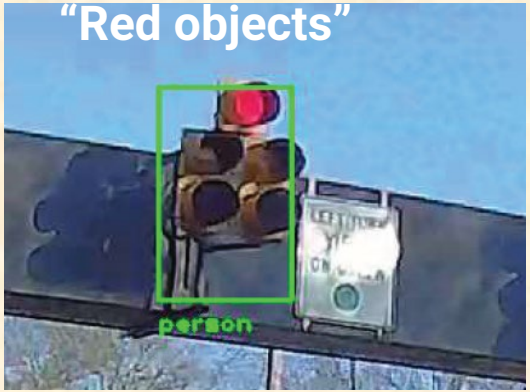
# A Few False Negatives Found:

EDGE CASE RESEARCH











**Baseline, un-augmented images with Mask-R CNN**
**// Your mileage may vary.**

# Human Intuition Isn't Enough

EDGE CASE RESEARCH

- **Many surprises aren't obvious to humans**
  - Unlikely to be included in human-designed tests
- **Example:**
  **High visibility clothing missed by perception**
- **Pilot study on real system:**
  - 82% recall of false negatives compared to ground truth

# Recalls are the Main NHTSA Hammer

Carnegie Mellon University

- ***Small*** **sampling NHTSA recalls (confirmed bugs)**
  - 17V-713: Engine does not reduce power / ESP software
  - 17V-686 *and MANY others*: Airbags disabled
  - 15V-569: Unexpected steering motion / loss of control
  - 15V-460 *and others*: Airbags deploy when they should not
  - 15V-145: Unattended vehicle starts engine→ carbon monoxide poisoning
  - 14V-370: Turns off headlights when driving
  - 14V-204: 1.5 seconds reverse while displaying Drive

*Voluntary Recalls:*

- 2018 hybrid engine stall at high speeds (https://bloom.bg/2y21T71)
- **2014 sudden unintended acceleration (https://goo.gl/R9zgL1)**

# Toyota "Unintended Acceleration" Has Killed 89

A 2005 Toyota Prius, which was in an accident, is seen at a police station in Harrison, New York, Wednesday, March 10, 2010. The driver of the Toyota Prius told police that the car accelerated on its own, then lurched down a driveway, across a road and into a stone wall. (AP Photo/Seth Wenig) / **AP PHOTO/SETH WENIG**

Unintended acceleration in Toyota vehicles may have been involved in the deaths of 89 people over the past decade, upgrading the number of deaths possibly linked to the massive recalls, the government said Tuesday.

The National Highway Traffic Safety Administration said that from 2000 to mid-May, it had received more than 6,200 complaints involving sudden acceleration in Toyota vehicles. The reports include 89 deaths and 57 injuries over the same period. Previously, 52 deaths had been suspected of being connected to the problem.

**May 25, 2010**

http://www.cbsnews.com/news/toyota-unintended-acceleration-has-killed-89/

# It's All Your Fault: The DOT Renders Its Verdict on Toyota's Unintended-Acceleration Scare

The final word on the Toyota unintended-acceleration mess.

*CSABA CSERE* JUN 9, 2011


From the June 2011 issue of *Car and Driver*

Carnegie Mellon University

- US DOT misrepresented the NASA report
- 2012: Class action settles
- 2013: Jury trial concluded it was the electronics
  - ~500 settlements
- 2014: $1.2B criminal fine
- My one hour talk on this: https://youtu.be/NCTf7wT5WR0 (Or search: Koopman Toyota UA )

https://www.caranddriver.com/features/a15125313/its-all-your-fault-the-dot-renders-its-verdict-on-toyotas-unintended-acceleration-scare-feature/

# "It's the Drivers' Fault"

Carnegie Mellon University

However, for most SAI, the most plausible cause of an open-throttle condition while attempting to brake is pedal misapplication, which is likely to be perceived as brake failure.

- Pollard & Sussman, 1989

- **"Most crashes are due to human error, therefore all unexplained crashes are due to human driver error"**
  - These statements trace back to this 1989 report
    - *Note: the reasoning is a logical fallacy*
  - US DOT reports fail to rule in software as a possible cause
- **Investigations:**
  - No mechanical cause found ➔ driver error
    - Compelling facts supporting human results in "unexplained"
  - Non-reproducible behavior ➔ driver error
    - "Pedal Misapplication" often blamed

2010


WIRED Operator Erro

JASON PAUR GEAR 03.12.10 12:15 PM

OPERATOR ERROR USUALLY THE CAUSE OF UNINTENDED ACCELERATION IN PAST INVESTIGATIONS

https://www.wired.com/2010/03/unintended-acceleration/

# Birth of the Pedal Misapplication Narrative

Carnegie Mellon University

- **Audi 5000: before full authority computer throttle control**
  - Public narrative: driver pedal mis-application & pedal placement

➔ *the same Pollard & Sussman report saying "pedal misapplication!"*

Among the principal conclusions were: 1) Some versions of Audi idle-stabilization system were prone to defects which resulted in excessive idle speeds and brief unanticipated accelerations of up to 0.3g. These accelerations could not be the sole cause of SAIs, but might have triggered some SAIs by startling the driver. 2) The pedal and seating arrangements of the Audi are significantly different from larger domestic cars. These differences may contribute to a higher incidence of pedal misapplication, especially for relatively unfamiliar drivers. 3) Brake failures are very unlikely and would be detectable after the event if they occurred.

Pollard & Sussman, 1989, DOT-TSC-NHTSA-88-4 Appendix H; 1983-85 Audi 5000

**Note: 0.3g is 0-to-60mph in 9.1 seconds; 1983 Audi 5000S 0-60 track time is 10.7 sec.**

https://www.zeroto60times.com/vehicle-make/audi-0-60-mph-times/

# Actual Pedal Misapplication Data

Carnegie Mellon University

- **Gas/Brake confusion 1 out of 997 (Pre-ETC data)**

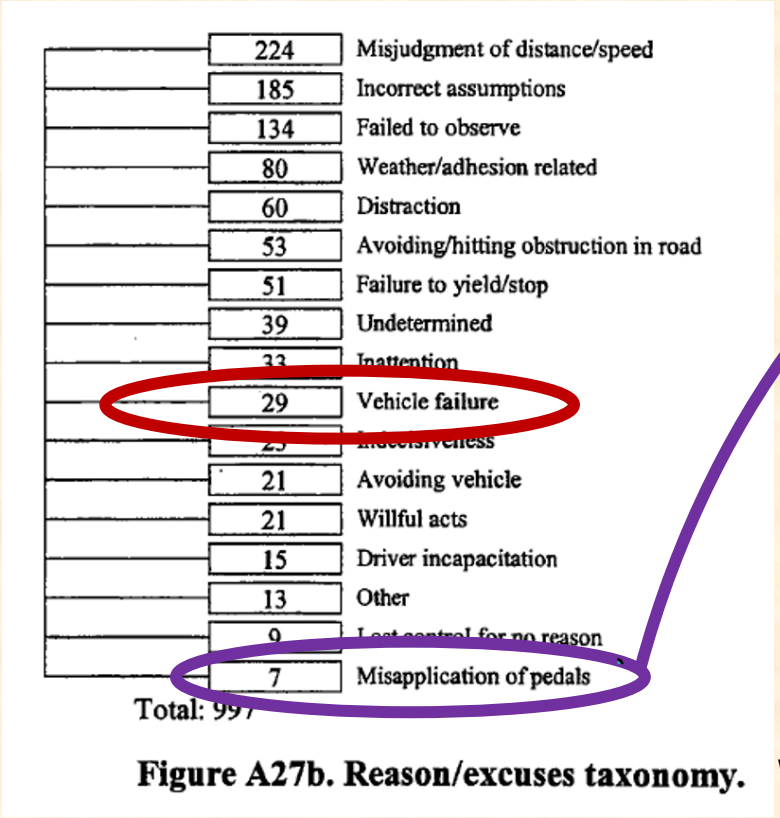
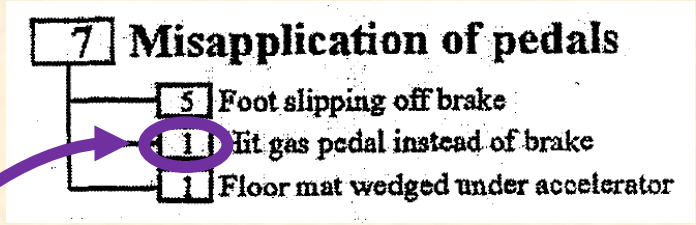



**Figure A27b. Reason/excuses taxonomy.**




- **Other data supports this**
- **Contradicting reports fail to take into account possibility of software defect**

Wierwille at al., FHWA-RD-02-003, **2002**

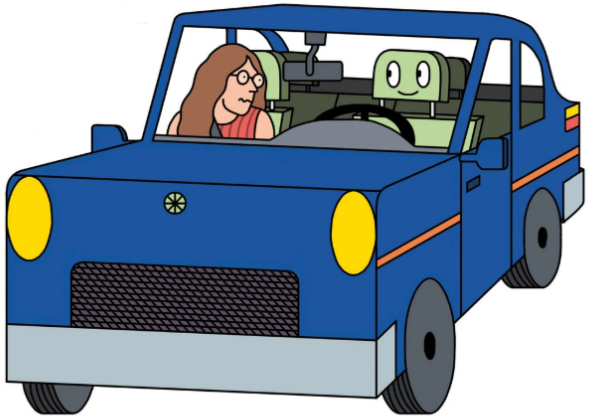
Opinion

https://www.nytimes.com/2017/10/24/opinion/self-driving-cars-safety.html

LETTERS

# Will Self-Driving Cars Improve Safety?

Oct. 24, 2017



Jan Buchczik

Carnegie Mellon University

- Won't drive drunk

...

- But won't be perfect
- Need safety standards
- Will self-certification work?

# The Day Automotive Safety Changed

EDGE CASE RESEARCH

## November 7, 2017

- **If there is no driver…**
  - Who do you blame when something goes wrong?
  - (Hint: not the driver)
- **Existing safety standards & practices are essential…**
  - (Car companies should actually follow them.)
  - But more is needed for autonomy

Waymo is first to put fully self-driving cars on US roads without a safety driver

*Going Level 4 in Arizona*

By Andrew J. Hawkins | @andyjayhawk | Nov 7, 2017, 11:00am EST

SHARE

https://bit.ly/2Vjjvrv

# UL 4600 Autonomous Vehicle Safety

EDGE CASE RESEARCH

- **Standardizes what to put in a safety case:**
  - Why do you think you are safe?
  - Where is the evidence to prove your reasoning is correct?
  - There are many known hazards:
    #DidYouThinkofThat?
- **Underwriters Laboratories / ANSI Standard**
  - Non-profit Standards Development Organization
  - Majority of starting point written by Edge Case Research
  - Improvements in response to hundreds of comments
    - Special thanks:
      Uma Ferrell, Frank Fratrik, Deborah Prince, Jason Smith

# System & Lifecycle Issues

EDGE CASE RESEARCH

- **Drivers do more than drive – there is no "captain of the ship"**
  - Ensure ready to operate
  - Mitigate equipment failures
- **Safety related lifecycle participants**
  - Inspection & maintenance accuracy
  - Supply chain faults
  - Field modifications & updates
- **Safety culture for all stakeholders**


Is it safe to drive now?

# Anticipated UL 4600 Technical Benefits

EDGE CASE RESEARCH

- **One year / ~ 300 pages**
- **Mandates safety case approach**
  - Why do you think you're safe?
  - What evidence support?
- **#DidYouThinkofThat? catalog**
  - Avoid missed hazards
  - Avoid pitfalls
  - Mechanism for industry to share hazards & lessons learned

EDGE CASE RESEARCH

HOME PRODUCTS & SERVICES COMPANY RESOURCES NEWS CONNECT

Carnegie Mellon University

Webinars

**TECHNICAL WEBINAR**


UL and Edge Case Research provide a technical discussion about UL 4600: Standard for Safety for the Evaluation of Autonomous Products. The webinar provides an overview of the document and how to participate in the review process.

**Download Slides for the Technical Webinar**

**Techical Webinar Q&A**

**GENERAL AUDIENCE WEBINAR**


UL and Edge Case Research provide an overview of UL 4600: Standard for Safety for the Evaluation of Autonomous Products. This presentation is for regulators, policymakers, media, and non-technical stakeholders. The webinar will discuss background, highlights, fit with ISO Standards, and other topics.

**Download Sides for the General Audience Webinar**

**General Audience Webinar Q&A**

**UL4600.com**

- **Youtube webinars**
- **Download full copy of draft UL 4600**

# UL 4600 Progress

EDGE CASE RESEARCH

- **Final comments on a minor revision pending**
  - 150+ non-voting stakeholders from 20 countries
  - Voting Standards Technical Panel members include:
    - Autonomy:  Uber, Argo, Aurora, Zenuity, Nissan NA, Locomation
    - Components: Infineon, Renesas, Intel/Mobileye
    - Insurance: Liberty Mutual, Munich RE, AXA XL
    - Government: US DOT, CPSC, PennDOT, MITRE, Oak Ridge NL
    - Universities: York, Nanyang, KTH, Waterloo, Beijing
    - Tools: Edge Case Research, ANSYS
    - Others: Center for Auto Safety, Intertek, UL LLC
- **Expect UL 4600 official issue in March 2020**

# Closing Thoughts

Carnegie Mellon University

- **Self-driving cars safer than humans?**
  - That sets the bar pretty high!
  - UL 4600 is a first draft of what it will take to get there
- **Fundamental shift in safety**
  - Perception/prediction novelty
  - No human Captain of the Ship
  - No human driver to blame when things go wrong

http://bit.ly/2MTbT8F (sign modified)